ATLAS CONF Note Draft

CONF-FTAG-2017-02

Version 0.1

Comments are due by: DD Month 2017

Supporting internal notes

ATL-COM-PHYS-2017-192: https://cds.cern.ch/record/2253746

Calibration of light-flavour jet b-tagging rates on ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV

A variety of algorithms have been developed to identify jets originating from b-quark hadronization within the ATLAS experiment at the Large Hadron Collider. We describe two measurements of the misidentification rate of jets containing no b- nor c-hadrons for the algorithm most commonly used in the LHC Run 2 ATLAS analyses. The measurements are performed in various ranges of jet transverse momenta and pseudorapidities based on proton-proton collision data collected at a centre-of-mass energy of $\sqrt{s} = 13$ TeV during the year 2015 and 2016. The first measurement is based on a data sample enriched in jets originating from light-flavour quark and gluon hadronization with the application of a dedicated algorithm reversing some of the criteria used in the nominal identification algorithm. The second measurement is based on a bottom-up approach where the underlying tracking variables in the simulation are adjusted to match the data. The effect is then propagated to the high-level observables relevant for b-identification. The results of both methods are found in good agreement and compared to the misidentification rate predicted by the nominal ATLAS simulation in order to calibrate it.

Document created on 2nd November 2017 using ATLAS LaTeX Version 04-02-00.

Analysis Team

[*email:* atlas-conf-ftag-2017-02-editors@cern.ch]

Valentina Cairo, Stefano Cali, Julian Glatzer, Emily Graham, Kristian Gregersen, David Jamin, Krisztian Peters, Matthias Saimpert, Balthasar Schachtner, Federico Sforza, Jonathan Shlomi

ATLAS CONF Note
CONF-FTAG-2017-02
2nd November 2017
Draft version 0.1

Calibration of light-flavour jet b-tagging rates on ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV

The ATLAS Collaboration

A variety of algorithms have been developed to identify jets originating from b-quark hadronization within the ATLAS experiment at the Large Hadron Collider. We describe two measurements of the misidentification rate of jets containing no b- nor c-hadrons for the algorithm most commonly used in the LHC Run 2 ATLAS analyses. The measurements are performed in various ranges of jet transverse momenta and pseudorapidities based on proton-proton collision data collected at a centre-of-mass energy of $\sqrt{s} = 13$ TeV during the year 2015 and 2016. The first measurement is based on a data sample enriched in jets originating from light-flavour quark and gluon hadronization with the application of a dedicated algorithm reversing some of the criteria used in the nominal identification algorithm. The second measurement is based on a bottom-up approach where the underlying tracking variables in the simulation are adjusted to match the data. The effect is then propagated to the high-level observables relevant for b-identification. The results of both methods are found in good agreement and compared to the misidentification rate predicted by the nominal ATLAS simulation in order to calibrate it.

Contents

1 Introduction 4

2 ATLAS detector 5

3 Data sample and event selection 6

4 Monte Carlo simulation 7

5 *b*-tagging algorithms 7

6 Track impact parameter modeling in the ATLAS simulation 7

7 Calibration with the negative tag method 7

8 Calibration with the adjusted Monte Carlo method 7

9 Comparison between the two methods 7

10 Conclusion 7

Appendix 9

Auxiliary material 12

1 Introduction

The identification of jets originating from the hadronization of a b-quark (b-tagging) is an important element of a number of prominent analyses performed with the ATLAS detector [1] at the Large Hadron Collider (LHC): measurements of standard model processes aiming to constrain the heavy-flavour (HF) parton density functions [2], studies of the top quark [3] and of the Higgs-boson [4, 5], and exploration of New Physics scenarios [6, 7].

The b-tagging of a jet relies on the property of the b-hadrons to have long lifetime τ ($\tau \sim 1.5$ ps, corresponding to a proper decay length of about $c\tau \sim 450\ \mu$m) and large mass, resulting in the production of tracks with non-zero impact parameters, secondary decay vertices and a large multiplicity of decay products inside the jet cone. These observables are reconstructed with the help of the charged-particle tracking capability of the ATLAS inner detector [8]. The information is then combined using a multi-variate algorithm able to enhance the discrimination of a jet containing b-hadrons (b-jet) with respect to a jet containing no b-hadrons but c-hadrons (c-jet) or a jet containing no b-hadrons nor c-hadrons (LF-jet, LF standing for light-flavour). Specific selections on the output weight distribution of a given b-tagging algorithm are called working points (WP) and defined as a function of the average efficiency of tagging a b-jet as measured in a $t\bar{t}$ simulated sample [9, 10]. The algorithm most commonly used in the Run 2 ATLAS analyses is called MV2. It assigns to each jet a b-tagging discriminant variable ranging from -1 to 1. LF-jets (b-jets) MV2 distribution peaks towards -1 (1) whereas the c-jets lie between the two. The

2016 version of MV2, which is the one considered in this document, is trained with a background sample including 7% of c-jets and 93% of LF-jets. It is denoted MV2c10 in the following.

The performance of a b-tagging algorithm is characterised by the probability of tagging a b-jet (ε_b) and the probabilities of mistakenly tagging as a b-jet a c-jet (ε_c) or a LF-jet (ε_l), referred to as "mistag rates" in the following. Ideally, Monte Carlo (MC) simulations including the various quark flavours could be used to evaluate the b-tagging performance. However, additional calibration is often needed to account for differences between data and simulation, originating for instance from an imperfect description of the geometry of the detector. In practice, each working point of the algorithm is calibrated as a function of the jet transverse momentum ($p_{\mathrm{T}}^{\mathrm{jet}}$) and absolute pseudorapidity ($|\eta^{\mathrm{jet}}|$).[1]

This document presents the measurements of the LF-jet mistag rate on ATLAS proton-proton collision data recorded at a center-of-mass energy of $\sqrt{s}$ = 13 TeV for the MV2c10 WP listed in Table 1 [11] using two methods giving consistent results, the negative tag and the adjusted MC. The negative tag method consists in measuring the LF-jet mistag rate from a high statistics data sample enriched in LF-jets with the application of a dedicated algorithm reversing some of the criteria used in the nominal identification algorithm. The adjusted-MC method is based on a bottom-up approach where the underlying tracking variables are adjusted to match the data and the effect is then propagated to the b-tagging observables. b-jet performance are described in separate references [12–14].

Table 1: b-tagging MV2c10 working points considered in this document. Each WP is defined by a cut value X on the MV2c10 output weight distribution (MV2c10 discrimnant > X, MV2c10 discriminant values ranging in $[-1, 1]$). The resulting b-tagging efficiency ($\varepsilon_b^{\mathrm{MC}}$) and c- and LF-jet rejection rates ($1/\varepsilon_c$, $1/\varepsilon_l$) as measured in a $t\bar{t}$ simulated sample are also shown.

WP	Cut value X	$\varepsilon_b^{\mathrm{MC}}$	c-jet rejection	LF-jet rejection
85%	0.18	85%	3	34
77%	0.65	77%	6	134
70%	0.82	70%	12	381
60%	0.93	60%	35	1539

2 ATLAS detector

The ATLAS experiment [1] at the LHC is a multi-purpose particle detector with a forward-backward symmetric cylindrical geometry and a near 4π coverage in solid angle. It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadron calorimeters, and a muon spectrometer.

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The transverse energy is defined as $p_{\mathrm{T}} = E/\cosh(\eta)$.

The inner tracking detector (ID) covers the pseudorapidity range $|\eta| < 2.5$. It consists of silicon pixel, silicon micro-strip, and transition radiation tracking detectors. Among them, the pixel detector is crucial for b-jet identification. A new inner pixel layer, the Insertable B-Layer [15, 16] (IBL), was added before the start of Run 2, at a mean sensor radius of 3.2 cm from the beam-line.

Outside the ID, the lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A hadron (iron/scintillator-tile) calorimeter covers the central pseudorapidity range ($|\eta| < 1.7$). The end-cap and forward regions are instrumented with LAr calorimeters for both EM and hadronic energy measurements up to $|\eta| = 4.9$. The muon spectrometer surrounds the calorimeters and is based on three large air-core toroid superconducting magnets with eight coils each. Its bending power is in the range from 2.0 to 7.5 T m. It includes a system of precision tracking chambers and fast detectors for triggering.

A two-level trigger system, using custom hardware followed by a software-based level, is used to reduce the event storage rate to a maximum of around 1 kHz.

3 Data sample and event selection

The proton-proton collision data sample recorded by the ATLAS detector during the year 2015 and 2016 is used. The LHC beams were operated with proton bunches organised in “bunch train”, with a bunch spacing of 25 ns. Only events taken during stable beam conditions and satisfying detector and data-quality requirements are considered.

The data used in the measurements were recorded using a suite of single jet triggers [17], requiring in the event at least one hadronic jet with sufficient transverse energy $p_{\mathrm{T}}^{\mathrm{jet}}$ and absolute pseudorapidity $|\eta^{\mathrm{jet}}| < 3.2$. Hadronic jets are reconstructed from clustered energy deposits [18] in the ATLAS calorimeter with the anti-k_t algorithm [19] and a parameter $R = 0.4$. Given the very high rates of such events at the LHC, only a fraction of events satisfying this requirement were recorded at low and medium $p_{\mathrm{T}}^{\mathrm{jet}}$ due to computational power and data storage limitations. In order to optimise the statistical power of the measurements, each $p_{\mathrm{T}}^{\mathrm{jet}}$ bin requires the trigger with lowest prescale (i.e. with highest integrated luminosity) that is more than 99.9% efficient in that range. The integrated luminosity of the data sample therefore depends on $p_{\mathrm{T}}^{\mathrm{jet}}$ and ranges from 0.02 pb^{-1} (20 GeV $< p_{\mathrm{T}}^{\mathrm{jet}} <$ 60 GeV) to 36.1 fb^{-1} ($p_{\mathrm{T}}^{\mathrm{jet}} >$ 500 GeV).

Events are required to have at least one reconstructed vertex with at least two associated well-reconstructed tracks. Futhermore, at least two jets reconstructed within the ATLAS inner detector pseudorapidity acceptance ($|\eta^{\mathrm{jet}}| < 2.5$) passing cleaning criteria [20], identified as coming from the primary hard interaction [21] and satisfying $p_{\mathrm{T}}^{\mathrm{jet}} >$ 20 GeV after final calibration [22] must be present. If more than two jets satisfy these criteria, the two jets with the highest transverse momenta are selected and the others are disregarded. A good angular separation between the two jets in the transverse plane ($\Delta\phi_{jj} > 2$ rad.) is also required in order to reject events with high transverse momentum jets originating from the hadronization of a gluon which split into two quarks ($g \rightarrow q\bar{q}$), more likely to contain c- and b-jets, or beam-induced background due to proton losses upstream of the interaction point [23].

4 Monte Carlo simulation

Samples of inclusive dijet events from strong interaction processes are generated with Pythia 8.186 [24] MC generator with the NNPDF 2.3 LO parton distribution functions (PDFs) [25]. This generator utilizes leading-order perturbative quantum chromodynamics (pQCD) matrix elements for $2 \rightarrow 2$ processes, along with a leading-logarithmic parton shower [26], an underlying event (UE) simulation with multiple parton interactions, and the Lund string model for hadronisation [27]. The parameters for the modelling of the interaction features not represented by the matrix element are provided by the A14 tune [28]. Alternative samples of inclusive dijet events from strong interaction processes are generated with HERWIG++ 2.7.1 MC generator [29] with the CTEQ6L1 LO PDFs [30] and the UEEE5 tune for the modelling of the interaction feature not represented by the matrix elements (parton shower, hadronization, underlying event).

Generated events are propagated through a full simulation of the ATLAS detector [31] based on Geant4 [32] that simulates the particle interactions with the detector material. All generated events are part of the MC15c campaign, which was processed with the release 20.7 of the ATLAS software. Hadronic showers are simulated with the FTFP BERT model [33]. Different pileup conditions as a function of the instantaneous luminosity are taken into account by overlaying simulated minimum bias events generated with Pythia8 onto the hard-scattering process and reweighting them according to the distribution of the mean number of interactions observed in data.

Residual differences between data and simulation regarding jet cleaning requirement efficiencies are at the few percent-level and corrected by means of event-wide scale factors applied to the simulated events. No corrections related to b-tagging performance are applied to MC simulation for the measurements.

5 *b*-tagging algorithms

6 Track impact parameter modeling in the ATLAS simulation

7 Calibration with the negative tag method

8 Calibration with the adjusted Monte Carlo method

9 Comparison between the two methods

10 Conclusion

Acknowledgements

The `atlaslatex` package contains the acknowledgements that were valid at the time of the release you are using. These can be found in the `acknowledgements` subdirectory. When your ATLAS paper or

PUB/CONF note is ready to be published, download the latest set of acknowledgements from:
https://twiki.cern.ch/twiki/bin/view/AtlasProtected/PubComAcknowledgements

Appendix

In a paper, an appendix is used for technical details that would otherwise disturb the flow of the paper. Such an appendix should be printed before the Bibliography.

References

[1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[2] M. Aaboud et al., *Measurement of differential cross sections of isolated-photon plus heavy-flavour jet production in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector*, (2017), arXiv: 1710.09560 [hep-ex].

[3] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s}$ = 7 and 8 TeV with the ATLAS detector*, Eur. Phys. J. C **74** (2014) 3109, arXiv: 1406.5375 [hep-ex].

[4] M. Aaboud et al., *Evidence for the $H \rightarrow b\bar{b}$ decay with the ATLAS detector*, (2017), arXiv: 1708.03299 [hep-ex].

[5] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, ATLAS-CONF-2017-076, 2017, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2017-076.

[6] M. Aaboud et al., *Search for a scalar partner of the top quark in the jets plus missing transverse momentum final state at $\sqrt{s}$=13 TeV with the ATLAS detector*, (2017), arXiv: 1709.04183 [hep-ex].

[7] ATLAS Collaboration, *Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the $q\bar{q}^{(\prime)}b\bar{b}$ final state in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, (2017), arXiv: 1707.06958 [hep-ex].

[8] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex].

[9] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, 2015, URL: https://cds.cern.ch/record/2037697.

[10] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, URL: https://cds.cern.ch/record/2160731.

[11] ATLAS Collaboration, *b-tagging calibration plots in light-flavour jets from 2015+2016 data using the negative-tag method*, FTAG-2017-002, 2017, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2017-002.

[12] ATLAS Collaboration, *b-tagging calibration plots using dileptonic tt events produced in pp collisions at $\sqrt{s}$ = 13 TeV and a combinatorial likelihood approach*, FTAG-2016-003, 2016, URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2016-003.

[13] ATLAS Collaboration, *b-tagging efficiency calibration using a tag-and-probe technique with opposite-sign, different-flavour dileptonic tt events produced in pp collisions at* $\sqrt{s} = 13$ *TeV*, FTAG-2017-001, 2017, URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2017-001.

[14] ATLAS Collaboration, *b-tagging efficiency calibration from 2015+2016 data on lepton+jets tt events*, FTAG-2017-003, 2017, URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2017-003.

[15] ATLAS Collaboration, 'ATLAS Insertable B-Layer Technical Design Report', tech. rep. CERN-LHCC-2010-013, CERN, 2010, URL: https://cds.cern.ch/record/1291633.

[16] ATLAS Collaboration, *Plots of b-tagging performance before and after the installation of the Insertable B-Layer*, FTAG-2017-005, 2017, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2017-005.

[17] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. C **77** (2017) 317, arXiv: 1611.09661 [hep-ex].

[18] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex].

[19] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph].

[20] ATLAS Collaboration, *Selection of jets produced in* 13 *TeV proton–proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: https://cds.cern.ch/record/2037702.

[21] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, URL: https://cds.cern.ch/record/1700870.

[22] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, (2017), arXiv: 1703.09665 [hep-ex].

[23] ATLAS Collaboration, *Characterisation and mitigation of beam-induced backgrounds observed in the ATLAS detector during the 2011 proton–proton run*, JINST **8** (2013) P07004, arXiv: 1303.0223 [hep-ex].

[24] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph].

[25] R. D. Ball et al., *Parton distributions with QED corrections*, Nucl. Phys. **B877** (2013) 290, arXiv: 1308.0598 [hep-ph].

[26] R. Corke and T. Sjostrand, *Improved Parton Showers at Large Transverse Momenta*, Eur. Phys. J. **C69** (2010) 1, arXiv: 1003.2384 [hep-ph].

[27] B. Andersson, G. Gustafson, G. Ingelman and T. Sjostrand, *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31.

[28] ATLAS Collaboration, *ATLAS Pythia8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419.

[29] J. Bellm et al., *Herwig++ 2.7 Release Note*, (2013), arXiv: 1310.6877 [hep-ph].

[30] J. Pumplin et al.,
New generation of parton distributions with uncertainties from global QCD analysis,
JHEP **07** (2002) 012, arXiv: hep-ph/0201195 [hep-ph].

[31] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823,
arXiv: 1005.4568 [hep-ex].

[32] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[33] G. Folger and J. P. Wellisch, *String parton models in GEANT4*,
eConf **C0303241** (2003) MOMT007, arXiv: nucl-th/0306007 [nucl-th].

Auxiliary material

In an ATLAS paper, auxiliary plots and tables that are supposed to be made public should be collected in an appendix that has the title 'Auxiliary material'. This appendix should be printed after the Bibliography. At the end of the paper approval procedure, this information should be split into a separate document – see `atlas-auxmat.tex`.